

## Linear Combinations of Means in Completely Randomized Designs

$C = k_1\mu_1 + k_2\mu_2 + \cdots + k_t\mu_t = \sum_{i=1}^t k_i\mu_i$  is a linear combination of the means  $\mu_1, \mu_2, \dots, \mu_t$

$C$  is estimated by  $\hat{c} = k_1\hat{\mu}_1 + k_2\hat{\mu}_2 + \cdots + k_t\hat{\mu}_t = \sum_{i=1}^t k_i\hat{\mu}_i = \sum_{i=1}^t k_i\bar{y}_i$ .  $\text{Mean}(\hat{c}) = C$

$C = k_1\mu_1 + k_2\mu_2 + \cdots + k_t\mu_t = \sum_{i=1}^t k_i\mu_i$  is a *contrast* if  $\sum_{i=1}^t k_i = k_1 + k_2 + \cdots + k_t = 0$ .

1. Suppose in our diet example that diets 1 and 3 are high in protein and diet 2 is lower in protein. Determine the values of  $k_1, k_2$ , and  $k_3$  that will give a contrast that compares the mean of the high protein diets to the mean of the lower protein diet.

2. In our diet example, the treatment means were  $\bar{y}_1 = 32, \bar{y}_2 = 25$ , and  $\bar{y}_3 = 30$ . Estimate the contrast defined in question 1.

The variance of a linear combination of means  $\hat{c} = \sum_{i=1}^t k_i\bar{y}_i$  is  $\text{Var}(\hat{c}) = \sum_{i=1}^t k_i^2 \left( \frac{\sigma^2}{r_i} \right) = \sigma^2 \sum_{i=1}^t \frac{k_i^2}{r_i}$

The standard error of a linear combination of means  $\hat{c} = \sum_{i=1}^t k_i\bar{y}_i$  is  $\text{SE}(\hat{c}) = \sqrt{\text{MSE} \sum_{i=1}^t \frac{k_i^2}{r_i}}$

3. In our diet example, there were 4 replications in each treatment group and the sum of squared errors was 78. Find the standard error of the contrast estimated in question 2.

Recall that  $\frac{\text{Estimate} - \text{Mean}(\text{Estimate})}{\text{SE}(\text{Estimate})}$  has (under certain conditions) a  $t$ -distribution with degrees of freedom equal to the degrees of freedom associated with the estimate of  $\sigma^2$  in  $\text{SE}(\text{Estimate})$ .

4. Was the mean of the high protein diets significantly different from the mean of the lower protein diet? Conduct a  $t$ -test to answer this question.

Recall that a general formula for a  $100(1 - \alpha)\%$  confidence interval is

$$\text{Estimate} \pm t_{\alpha/2, d.f.} \text{SE}(\text{Estimate}),$$

where  $t_{\alpha/2, d.f.}$  is the value that is matched or exceeded with probability  $\alpha/2$  by a random variable with a  $t$ -distribution with degrees of freedom= $d.f.$  The degrees of freedom matches the degrees of freedom associated with the estimate of  $\sigma^2$  in  $\text{SE}(\text{Estimate})$ .

5. Find a 95% confidence interval for the mean of the high protein weight gain means minus the lower protein weight gain mean.

The sum of squares associated with a contrast  $C$  is

$$SSC = \frac{\hat{c}^2}{\sum_{i=1}^t k_i^2/r_i} = \frac{\left(\sum_{i=1}^t k_i \bar{y}_i\right)^2}{\sum_{i=1}^t k_i^2/r_i} = \frac{(k_1 \bar{y}_1 + k_2 \bar{y}_2 + \cdots + k_t \bar{y}_t)^2}{k_1^2/r_1 + k_2^2/r_2 + \cdots + k_t^2/r_t}$$

The contrast  $\hat{c}$  has 1 degree of freedom because it involves only 1 linear combination of means. Thus the mean square for the contrast is  $MSC = SSC/1$ .

An  $F$ -test can be used to test  $H_0 : C = 0$  vs.  $H_A : C \neq 0$ . The  $F$ -statistic is  $MSC/MSE$ . When  $H_0$  is true, the statistic has an  $F$ -distribution with 1 degree of freedom in the numerator and denominator degrees of freedom matching the degrees of freedom for error.

6. Conduct an  $F$ -test to determine if the mean of the high protein diets is the same as the mean of the lower protein diet.

7. How does your test compare with the  $t$ -test that you conducted in question 4?

Two contrasts  $C = \sum_{i=1}^t k_i \mu_i$  and  $D = \sum_{i=1}^t \ell_i \mu_i$  are *orthogonal* if  $\sum_{i=1}^t \frac{k_i \ell_i}{r_i} = 0$ .

Write down the coefficients of a contrast for comparing the means of the two high protein diets. Is this contrast orthogonal to the contrast that compares the mean of the high protein diets to the lower protein diet?

Suppose  $C$  and  $D$  are orthogonal contrasts, and we wish to test

$$H_0 : C = 0 \text{ and } D = 0 \text{ vs. } H_A : \text{ either } C \neq 0, D \neq 0, \text{ or both.}$$

We may conduct an  $F$ -test where the  $F$ -statistic is given by

$$F = \frac{(SSC + SSD)/2}{MSE}.$$

When the null hypothesis is true, this statistic has an  $F$ -distribution with 2 degrees of freedom in the numerator and denominator degrees of freedom matching the degrees of freedom for error.

Conduct one  $F$ -test of the null hypothesis that says that the mean of the high protein diets is the same as the mean of the lower protein diet AND the means of the two high protein diets are the same.